#5

04-19-04 P02:57 IN

Evaluation of the Boston Study of Effectiveness of Soil Abatement in Reducing Children's Blood Lead, with Particular Emphasis Upon the EPA (1996) Reevaluation

Kenny Crump, Ph.D.
ICF Kaiser
Ruston, Louisiana 712170
Tel: 318-242-5019
Fax: 318-255-4960
E-mail: kcrump@iAmerica.net

March 13, 1997

Executive Summary

The Three Cities Lead Study investigated the effect of abatement of lead in soil upon blood lead levels in children living in Boston (EPA, 1993a; Weitzman *et al.*, 1993; Aschengrau *et al.*, 1994), Baltimore (EPA, 1993b) and Cincinnati (EPA, 1993c). This review focuses upon the Boston Study and, in particular, upon the EPA (1996) reanalysis of these data.

In the Boston Study, 152 children less than 4 years of age and with blood lead levels between 7 and 24 μg/dL were enrolled and randomly assigned to one of three treatment groups. These groups received the following interventions during Phase I of the study: interior loose paint removal (Group B); interior loose paint removal and interior dust abatement (Group A); interior loose paint removal, interior dust abatement, and soil lead abatement (Group S). During Phase II, the soil of Groups A and B was abated and about one-half of the properties in all three groups experienced interior paint deleading. Since there was no appropriate control group for the soil remediation performed during Phase II, and since during Phase I Groups B and S differed with respect to both soil and dust abatement, statistically valid conclusions regarding the effect of soil lead abatement must be based on comparisons between Groups A and S during Phase I.

EPA's reanalysis of the data from the Boston Study (EPA, 1996) represents a very substantial effort on the part of the Agency in terms of the statistical

analyses conducted. Unfortunately, this effort is beset by a number of problems, including poor selection of statistical methods, failure to adequately examine the role of potential confounding variables, highly selective interpretation of results, and listing as conclusions statements that are not supported by the analyses. These problems are so serious that we think that it would be dangerous to draw any conclusions from the EPA analysis regarding the effect of soil abatement upon blood lead levels.

On the one hand, EPA presents results of an enormous amount of work based on structural equation modelling, which EPA described as "an exploratory model-building activity" which "does not readily lend itself to hypothesis testing." It is highly questionable as to whether the structural equation modelling effort successfully fulfilled even this limited purpose, as the models appeared to be unstable, and many of them apparently failed to converge. As a result, this extensive exercise proved to be more informative regarding structural equation modelling *per se* than it did regarding the effect of soil abatement on children's blood lead levels.

On the other hand, EPA's statistical analyses included only a very limited investigation into the potential role of confounding factors in the Boston data. Of 23 potential confounding factors considered by Weitzman *et al.* (1993), EPA considered only three (race, age and gender) in any of their analyses. Moreover, these three factors were investigated mainly individually rather than jointly, and then only in a rudimentary way by analyzing subsets of the data defined by these factors.

A number of EPA's analyses do not take into account important features in the experimental design. For example, EPA's analysis of Phase II data takes no account of the fact that about one-half of the premises had their interior paint deleaded. Another example is an analysis by EPA that showed a significant correlation between post-abatement soil lead and post-abatement house dust lead. However, this analysis did not control for treatment, and consequently this correlation is likely due to confounding with treatment.

The most troubling feature of the EPA (1996) analysis of the Boston Study is that a number of EPA's conclusions are either not supported by their analyses, or are apt to mislead a reader regarding the effect of soil lead. Examples of such conclusions are as follows:

"Blood lead levels were reduced by approximately 1.86 μg/dL at 10 mo after soil abatement."

Actually this value was the difference in reductions in blood lead between Groups S and B, which, as noted above, cannot be attributed to soil abatement.

"The reductions of lead in both soil and house dust persisted for at least two years."

Reductions in house dust lead were essentially the same in all three study groups, and consequently cannot be attributed to soil abatement, which did not occur in two of the three groups in Phase I.

"Soil was clearly a part of the exposure pathway to the child, contributing significantly to house dust lead"

The analysis that EPA apparently based this conclusion upon regressed post-abatement house dust levels upon post-abatement soil levels, without controlling for treatment effect. Consequently, the positive correlation between soil lead and dust lead is likely to have been a result of confounding with treatment effect. Analyses by EPA that controlled for treatment did not obtain this result. Actually, analyses of pre-abatement soil and dust levels are more appropriate for evaluating this question. However, EPA's analyses of pre-abatement levels tended to show negative correlations between soil

and dust lead levels, and EPA ignored these negative correlations in its discussion and conclusions.

"Additional reductions in blood lead of about 2.0 μg/dL (relative to non-abated) were observed at 22 mo post-abatement for children in houses where the soil lead was abated and the interior house dust lead was consequently reduced and remained low."

It is not clear what the phrase "relative to non-abated" refers to, since by 22 months the soil of all the properties in the study had been abated. It is also not clear what EPA means by "additional reductions", since the median blood lead level of children in Group S (the only group observed for 22 months following soil abatement) did not change between 7 and 22 months post-abatement. Finally, "consequently reduced" seems to imply cause and effect, whereas, as noted above, it cannot be concluded from EPA's analyses that soil abatement reduced interior dust lead levels.

Another very troubling feature of EPA's reanalysis of data from the Boston Study is its selective analyses and interpretations. To cite one example, the result featured most prominently by EPA from its extensive structural equation modelling effort was an estimate of the effect of soil abatement on blood lead levels. This estimate was derived by EPA by modifying one of its 32 structural equation models by fixing one parameter at the value estimated in another model — a model that did not even include an effect of soil abatement on blood lead. We can think of no reason why EPA would emphasize the result of this particular analysis, other than the fact that it produced results that were more similar to the estimate obtained by Weitzman *et al.* (1993) than any of the others EPA obtained from its structural equation models.

Reading of the EPA report is complicated by numerous editing problems, including missing or poorly documented tables, tables with unexplained missing

entries, tables whose entries do not conform to the models described in the text, incorrect description of models, and omission of important modelling details. These problems are detailed subsequently in this evaluation.

In view of the many serious problems with the EPA analyses, we think it would be inappropriate to draw any conclusions from them regarding the effect of soil abatement upon blood lead levels beyond the findings reported from Phase I data by the original investigators (Weitzman *et al.* 1993). Since, curiously, the Weitzman *et al.* (1993) report was not cited at all in the EPA (1996) reanalysis, its conclusion is repeated here:

> "These results demonstrate that lead-contaminated soil contributes to the lead burden of urban soil and that the abatement of lead-contaminated soil around homes results in a modest decline in blood lead levels. The magnitude of reduction in blood lead levels observed, however, suggests that lead-contaminated soil abatement is not likely to be a useful clinical intervention for the majority of urban children in the United States with low-level lead exposure." (Weitzman *et al.*, 1993)

We conducted an independent non-parametric analysis of the effect of soil abatement upon blood lead that differed from the analysis performed by Weitzman *et al.* (1993) in two respects. First, this non-parametric analysis did not require the assumption that blood lead levels were normally distributed. Second, the analysis took into account the fact that premises rather than children were randomly assigned to study groups. When controlling jointly for pre-abatement blood lead, age, gender, race/ethnicity and socioeconomic status, this analysis estimated that abatement of soil containing an average of 2300 ppm lead resulted in a statistically significant average reduction in blood lead of 1.37 μg/dL, compared to the value of 1.28 μg/dL obtained by Weitzman *et al.* (1993). However, care must be taken when extrapolating this estimate to other situations and localities, as there are a number of reasons why this approach would tend to overestimate the effect of soil

abatement in general. First, the median soil lead level in Boston was in excess of 2000 ppm, which appears to be higher than in most urban settings. Abatement in urban areas with lower soil lead levels would likely result in smaller reductions in blood lead levels than were observed in the Boston Study. This is borne out by the Cincinnati and Baltimore studies, where soil lead levels were lower and no benefit of soil abatement in reducing blood lead levels was identified. Second, only children whose blood lead levels were between 7 and 24 μg/dL were included in the Boston Study, and it appears that the vast majority of excluded children had blood lead levels below 7 μg/dL. The reduction in blood lead resulting from soil abatement would likely be lower in such children. Third, both the Weitzman *et al.* analysis and, for consistency, our analysis omitted two children from Group S as outliers. Weitzman *et al.* indicated that the increases in blood lead in these two children were likely due to exposure to leaded paint at another site. However, exposures of this type occur in real life situations. When we included these two children included in the analysis, the effect of abatement was estimated as –0.17 μg/dL; i.e., there is an estimated net increase in blood lead level in the soil abatement group (Group S) relative to the control group (Group A). Fourth, the reductions in blood lead observed in the Boston Study were in groups that also experienced paint stabilization and dust abatement. Consequently, the Boston Study does not address the efficacy of soil abatement by itself, but only in the presence of these other interventions.

Background

During the past 25 years, evidence has mounted that developmental effects can occur in children from environmental exposures to lead. Potential sources of lead exposure in children include lead-based paint and lead contaminated soil. In 1986, EPA was mandated under the Superfund Amendments and Reauthorization Act (SARA) to conduct soil abatement projects in up to three U.S. cities. In response, EPA undertook the "Three Cities Lead-Study," the purpose of which was

to determine whether abatement of lead in soil could reduce lead in the blood of inner city children.

The Three Cities Lead Study was composed of individual studies of inner city children in Boston (EPA, 1993a; Weitzman, 1993; Aschengrau *et al.*, 1994), Baltimore (EPA, 1993b), and Cincinnati (EPA, 1993c). Although the protocols of the three studies differed, all involved identification of a study population of children likely to be exposed to neighborhood soil containing high levels of lead, placement of children into groups that were subject to different types of soil abatement, measurement of pre- and post-abatement levels of lead in soil, household dust, and children's blood, and determination of children's activity patterns, eating habits, family activities and socioeconomic status.

Based on Phase I of the Boston Study, Weitzman *et al.* (1993) concluded that "abatement of lead-contaminated soil around homes results in a modest decline in blood lead levels." Following the completion of Phase II of the Boston Study, Aschengrau *et al.* (1994) concluded that "The combined results from both phases suggest that a soil lead reduction of 2060 ppm is associated with a 2.25 to 2.70 μg/dL decline in blood lead levels." In contrast to the Boston Study, neither the Baltimore Study (EPA, 1993b) nor the Cincinnati Study (EPA, 1993c) found that removing lead-contaminated soil reduced blood lead levels. EPA (1996) conducted an integrated assessment of data from all three studies.

ICF Kaiser was asked by the law firm of Seeger, Potter, Richardson, Luxton, Joselow & Brooks to review the previous work on the Three Cities Lead Study and to independently assess the evidence from this study for an effect of soil abatement on children's blood lead levels. This report, which is in response to that request, focuses on the Boston Study.

Description of the Boston Study

The study population was drawn from an area in Boston where the surface soil lead level averaged more than 3000 ppm (parts per million). To be included in the study a child's blood lead had to be between 7 μg/dL (micrograms per deciliter)

and 24 μg/dL, and their homes and families had to satisfy certain criteria, including criteria related to age (not more than 4 years of age), presence of chipping or peeling paint, accessibility to a yard composed of dirt or grass, and average surface soil lead level (≥ 1500 ppm) (Weitzman *et al.*, 1993). Children selected as participants were randomly assigned by premise to one of three study groups (EPA, 1993a). This random assignment resulted in 54 children in Study Group S, 51 in Comparison Group A, and 47 in Comparison Group B. The following lead abatements were performed in the three groups at the beginning of Phase I of the study:

Group B (called "BOS P-S" by EPA, 1996) —	Interior paint stabilization
Group A (called "BOS PI-S" by EPA, 1996) —	Interior paint stabilization and interior dust abatement
Group S (called "BOS SPI" by EPA, 1996) —	Interior paint stabilization, interior dust abatement and soil abatement

During Phase I soil lead, household dust lead and blood lead samples were collected from participants before abatement began (Round 1). Blood and household dust were also sampled at approximately 6 months post-abatement (Round 2) and 11 months post-abatement (Round 3). Soil lead was also sampled in Round 3.

At the beginning of Phase II (Aschengrau *et al.*, 1994), soil was abated at the premises of children in Groups A and B who remained in the study. In addition, lead paint abatement was offered to all remaining participants, and this offer was accepted by 38% of the participants. These abatements occurred approximately one year after the original abatement activities during Phase I. Blood lead samples were then collected approximately 10 months following these abatements (Round 4).

Comments on issues common to different analyses of the Boston Study

Before discussing the three analyses of the Boston data (Weitzman *et al.*, 1993; Aschengrau *et al.*, 1994; EPA, 1996) in detail, we first discuss some aspects that are common to two or more of these analyses.

Selection of control group: As described above, Groups S and A in the Boston Study experienced paint stabilization and dust abatement, with Group S experiencing soil abatement in addition. Since the treatments applied to Groups S and A differ only with respect to soil abatement, differences in blood lead between Groups S and A can properly be attributed to soil abatement (assuming other confounding factors are not present). However, whereas Group S had both dust and soil abatement, Group B experienced neither of these interventions. Thus, in these two groups soil abatement is completely confounded with dust abatement. As a result, differences in blood lead between Groups S and B cannot be attributed solely to soil abatement.

Weitzman *et al.* (1993) did not ascribe differences between Groups S and B to soil abatement, as evidenced by their conclusion that "Soil abatement alone was associated with a decline in blood levels of ... 0.8 to 1.6 μg/dL" (which summarized values obtained from comparisons of Groups S and A) and "soil and interior dust abatement combined was associated with a decline of ... 1.2 to 1.6 μg/dL (which summarized values obtained in comparisons of Groups S and B)" [emphasis added].

However, Aschengrau *et al.* (1994) did use comparisons between Groups S and B to reach conclusions about the efficacy of soil abatement alone. In particular, they concluded that "combined results from both phases suggest that a soil lead reduction of 2060 ppm is associated with a 2.25 to 2.70 μg/dL decline in blood lead levels." The specific method used by Aschengrau *et al.* to arrive at this range is fairly involved, and will be reviewed in more detail in a following section. However, the range is based partially upon comparison of blood levels in Group S

with those in Group B during Phase I. Consequently, it is not appropriate to ascribe this decline in blood lead to soil abatement alone.

Although the tables in EPA (1996) generally contain both comparisons between Groups S and A and between Groups S and B, EPA emphasizes comparisons between Groups S and B, and generally attributes the difference in blood lead between these two groups entirely to soil abatement. This interpretation is evident in the following quotes from EPA (1996):

> "Blood levels were reduced by approximately 1.86 μg/dL [the value EPA obtained in a comparison of Groups S and B] at 10 months after soil lead abatement." (EPA, 1996, page 1-19 and 6-4)

> "The Boston study shows very clear evidence of an effect of soil lead abatement in reducing blood lead in children currently residing in lead-contaminated housing. The effect was detected in the whole group of children that received soil abatement, amounting to 1.9 μg/dL [round up of value of 1.86 EPA obtained in a comparison of Groups S and B] or about 17% on average, ..." (EPA, 1996, page 5-143)

EPA refers to Group B as the "control group" for Group S, and refers to differences between Groups S and B as the "effect size of soil abatement" (EPA, 1996, page 5-82). EPA stated that interior dust abatement was carried out "to enhance the impact of soil abatement."

At one point EPA does appear to acknowledge that differences between Groups S and B could be due to dust abatement rather than soil abatement:

> "Blood lead reduction of about 1.9 μg/dL associated with soil and <u>interior dust abatement</u> occurred during Phase I of the Boston study" [emphasis added] (EPA, 1996, page 5-150)

Since effects of soil and dust abatement are confounded in comparisons of Groups S and B, effects attributed to soil abatement alone must be derived from comparisons of Groups S and A. However, care must be taken even in these comparisons. Since both groups experienced interior dust abatement and interior paint stabilization, differences between Groups S and A represent (assuming other confounding factors are not responsible) differences in blood lead from soil abatement in the presence of paint stabilization and dust abatement. The same effect of soil abatement upon blood lead levels might not have occurred without simultaneous interior dust abatement and loose paint stabilization.

Elimination of outliers: In both Phase I and Phase II of the Boston Study, blood lead measurements in certain children were defined to be "outliers", and those measurements were eliminated from all analyses. Weitzman *et al.* identified as outliers blood lead measurements of 35 and 43 μg/dL found in two Group S siblings during Round 3. Weitzman *et al.* stated that each of these measurements was more than three standard deviations above the overall mean of Round 3 blood lead measurements, and noted that these particular children were exposed to leaded paint at another site that was being renovated. These two measurements were also eliminated from EPA's analyses. Aschengrau *et al.* additionally eliminated from their analyses a Group B child whose Round 4 blood lead level was 23 μg/dL. No other reason for this exclusion was provided and apparently EPA did not exclude this child in its analyses. Each of these three exclusions caused the estimate of the effect of soil remediation to increase.

For a statistical test to remain valid after elimination of outliers, the decision as to which outcomes to eliminate must not depend upon which treatment groups the outcomes are from. Consequently, although it not absolutely necessary to set up the decision rule before examining the data, such decisions should be made without knowledge of the group assignments. If an *a priori* decision rule is selected, that decision rule should not be defined in terms of the group assignments. For example, it would not have been valid for Weitzman *et al.* to

have decided to eliminate children from a group whose blood lead levels were more than three standard deviations from the mean for that group. With this decision rule, whether an outcome was eliminated could depend upon the treatment assignments, which would violate the basic assumption that all outcomes are equally likely under the null hypothesis.

Although the decision by Weitzman *et al.* to eliminate the two outliers apparently was not made without knowledge of the group assignments, the important thing is whether or not the same decision would have been made if the outcomes in question had been in a different group. Although there is no way to know for sure, their definition of an outlier is a reasonable one. The situation with regard to the child eliminated by Aschengrau *et al.* is not so clear, since the criterion used to identify this outlier is not provided.

Although exclusion of outliers does not necessarily invalidate a statistical analysis, it does affect how the results of analysis should be interpreted and the extent to which they can be generalized. Strictly speaking, the effect estimate obtained after elimination of the outliers applies only to the restricted sample. Thus, EPA's conclusion that "Blood lead levels were reduced by approximately 1.9 μg/dL at 10 months after soil abatement" (EPA, 1996, page 6-4) applies only to the restricted sample. Also, it cannot be assumed that it applies to children who do not experience lead poisoning since, for example, it is possible that some of the decreases in blood lead seen in the soil abatement group were caused by an earlier poisoning episode, rather than being due to intervention.

Description and critique of the Weitzman *et al.* (1993) analysis of Phase I data from the Boston Study

Weitzman *et al.* (1993) is the original investigators' published report of the results of Phase I of the Boston Study. Curiously, this key report was not referred to in the EPA (1996) review and reanalysis.

Weitzman *et al.* (1993) applied a standard analysis of covariance regression model to the Phase I data, using the untransformed post-abatement blood lead

levels as the dependent variable. In their base model, pre-abatement blood lead was used as an explanatory variable, along with indicator variables that specified abatement group. In other analyses, the base model was augmented by adding variables individually that controlled for a number of potential confounding variables, including age, sex, race/ethnicity, socioeconomic status, and 19 variables related to a child's behavior and home environment.

Based on their base model, Weitzman *et al.* found that post-abatement blood lead was significantly lower in Group S than in either Group A ($p = 0.02$) or Group B ($p = 0.01$). This result became non-significant when control for race/ethnicity or lead paint was added to the base model, but remained significant, or nearly so ($p \leq 0.06$), when the remaining variables were added individually.

Although the regression model used by Weitzman *et al.* is a reasonable one, it does have two potential limitations. First of all, it assumed that post-abatement blood lead levels are normally distributed, which does not appear to be the case. Second, it did not account for the fact that some children lived in the same household and consequently were likely to respond more similarly than children from different households. In a later section of the present report, an analysis is presented that is similar to that of Weitzman *et al.*, but which is not subject to these two drawbacks.

Since the Weitzman *et al.* (1993) report was not cited in the EPA (1996) reanalysis, its conclusion is repeated here:

> "These results demonstrate that lead-contaminated soil contributes to the lead burden of urban soil and that the abatement of lead-contaminated soil around homes results in a modest decline in blood lead levels. The magnitude of reduction in blood lead levels observed, however, suggests that lead-contaminated soil abatement is not likely to be a useful clinical intervention for the majority of urban children in the United States with low-level lead exposure." (Weitzman *et al.*, 1993)

Description and critique of the Aschengrau *et al.* (1994) analysis of Phase II data from the Boston Study

Aschengrau *et al.* (1994) report on results of the Phase II follow-up of the Boston cohort. In this phase, soil was abated from premises in Groups A and B. Although lead paint was abated in some homes in all three groups at the beginning of Phase II, children living in these homes were not included in the analysis by Aschengrau *et al.* Blood lead was sampled in all three groups prior to Phase II intervention (Round 3, at end of Phase I) and at Round 4, about one year following the Phase II intervention.

Aschengrau *et al.* concluded that the magnitude of the blood lead decline related to soil abatement was greater in Phase II than in Phase I, and the combined results from both phases suggested that a soil lead reduction of 2060 ppm was associated with a 2.25–2.70 μg/dL decline in blood lead levels. These declines were calculated by computing weighted averages of the decline in Phase I among Group S and the decline in Phase II among Groups A and B combined, omitting children who also had paint abatement during Phase II when estimating the effect of Phase II soil abatement. Two different weighted averages (2.89 and 3.34 μg/dL) were obtained by either including or excluding Phase I data from Group S children who had paint abatement in Phase II. These weighted averages were reduced by the mean blood lead decline (0.64 μg/dL) in Groups A and B during Phase I (considering only the 31 children in Groups A and B who did not have their paint deleaded during Phase II), which resulted in the final estimated declines related to soil lead abatement of 2.25 and 2.70 μg/dL.

These estimated declines can be represented as weighted averages of declines estimated for Group S in Phase I and Groups A and B in Phase II. For example, consider the figure of 2.70 μg/dL, based only upon children from Group S whose household paint was not abated during Phase II. Aschengrau *et al.* calculated a mean decrease in blood lead in Group S during Phase I of 3.03 μg/dL, which when reduced by the average decline of 0.64 μg/dL in Groups A and B during Phase I (based on only the 31 children in Groups A and B who did not have

their paint deleaded during Phase II), results in an estimated decline in Group S of 3.03 – 0.64 = 2.37 μg/dL. Similarly, Aschengrau *et al.* calculated a mean decrease in blood lead during Phase II in Groups A and B of 3.63 μg/dL, which, when reduced by the same average decline of 0.64 μg/dL in Groups A and B during Phase I, results in an estimated decline in Group A and B during Phase II of 3.63 – 0.64 = 2.99 μg/dL. The overall value of 2.70 μg/dL represents the following weighted average:

$$\frac{2.37 \times 29 + 2.99 \times 31}{29 + 31} = 2.70 \text{ μg/dl.}$$

The decline in blood lead related to soil abatement estimated from the Phase I study (2.37 μg/dL) is based upon only 29 of 52 available children from Group S and only 31 of 98 available children from Groups A and B combined. Furthermore, initial blood lead was not controlled for by Aschengrau *et al.*, and, as noted earlier, Group B is not an appropriate control group for Group S. Weitzman *et al.* (1993) estimated a blood lead decline related to soil abatement in Phase I of the Boston study of 1.28 μg/dL based upon all the data, controlling for initial blood lead level, and using Group A as the control. Clearly, the estimate of 1.28 μg/dL is a superior estimate of the effect of soil abatement upon blood lead from the Phase I Study than the value of 2.37 μg/dL assumed by Aschengrau *et al.*

The decline in blood lead of 3.63 μg/dL obtained by Aschengrau *et al.* for Phase II was based upon a combined group of children from Groups A and B for which there was no adequate control group. Although Aschengrau *et al.* attempted to use the experience of Groups A and B during Phase I as controls, this approach was not satisfactory (Aschengrau *et al.* refer to it as "less than ideal"), since there were a number of factors that this approach may not have adequately controlled for. All of the families received counseling about lead, including educational materials on the effects of lead and methods for reducing children's

lead exposure, and knowledge of the soil lead levels on their premises and blood lead levels of their children. As part of a community relations strategy, a Community Advisory Committee met regularly, and study staff remained in close contact with participating families. The blood lead reductions observed in both Phase I and Phase II may have been partially due to this educational program and this effect could have been different in the two Phases. As EPA (1996, page 1-30) noted, "All of these studies may have initiated behavioral changes from the moment of recruitment simply by informing parents and caretakers of potential lead hazards."

This does not present such a problem in interpreting Phase I results, assuming that both the treatment group (Group S) and control group (Group A) were equally impacted by this program. However, it presents a much more serious problem in Phase II, where there is no adequate control group. Blood lead levels have been decreasing generally (Pirkle *et al.*, 1994), and the reductions may partially reflect this general decline. Also, almost all of the children were more than 27 months of age during Phase II, which is approaching an age range in which blood lead levels generally fall. Furthermore, Aschengrau *et al.*'s estimate of 3.63 µg/dL was based upon only 31 of the initial 98 children in Groups A and B, and these 31 were self-selected, since they were all from families that refused paint deleading at the beginning of Phase II. What was really needed, but was unavailable, was a group of children who, except for soil abatement, had been treated exactly the same as the children whose soil was abated during Phase II. Since there was no such control group available for the Phase II soil abatement, no statistically valid conclusions can be drawn regarding the effect of soil lead abatement upon blood lead from Phase II of the Boston Study.

Description and Critique of the EPA (1996) Integrated Report

EPA performed an "integrated assessment" of the longitudinal studies that comprise the Three Cities Lead Study. This assessment involved creating a common data base of data from the three studies and conducting additional

analyses of data from the individual studies. No analyses were conducted that combined data from different cities (meta analyses). As described earlier, the present review concentrates upon the EPA reanalysis of data from the Boston Study.

EPA's intentions in this study were both worthy and ambitious. It is obvious that a great deal of effort went into EPA's analyses and into the preparations of the common data base and the report. Even though this critique focused only on EPA's work with the Boston Study, it proved to be quite an undertaking to review the large body of analyses of the Boston data presented by EPA — an undertaking which is probably still best described as incomplete. Therefore, EPA is to be commended for good intentions and good-faith effort. However, it is clear from reading the report that the available time and resources were not adequate to carry out EPA's ambitious undertaking.

On the one hand, EPA presents an enormous amount of work with structural equation modelling, which EPA described as "an exploratory model-building activity" whose purpose was to "elucidate pathways for environmental lead exposure from source to child" and which "does not readily lend itself to hypothesis testing" (EPA, 1996, page 5-68). It is questionable as to whether the structural equation modelling effort successfully fulfilled even this limited purpose, as the models appeared to be unstable, and many of them apparently failed to converge. As a result, this exercise proved to be more informative regarding structural equation modelling than it did regarding the effect of soil abatement on children's blood lead levels.

On the other hand, EPA's extensive modelling included only a very limited investigation into the potential role of confounding factors in the Boston data. Of 23 potential confounding factors considered by Weitzman *et al.* (1993), EPA considered only three (race, age and gender) in all of their analyses. Moreover, these three factors were investigated only in a rudimentary way by analyzing subsets of the data defined by these factors. Moreover, in almost all cases these factors were investigated one at a time and not as a group. Socio-economic status

– potentially an important confounding factor – was not even included in EPA's data base.

EPA's analyses of Phase II of the Boston Study took no account of the fact that almost half of the residences received lead paint abatement at the beginning of Phase II. In fact, there is no mention of this paint abatement activity in EPA's report. This omission, coupled with the fact that there was no adequate control group for Phase II soil abatement, indicates that it would be unwise to draw any conclusions from EPA's analyses of Phase II data regarding the effect of soil abatement upon blood lead levels.

Several of EPA's key conclusions are not linked by EPA to any particular analysis, and do not appear to be supported by EPA's analyses. These unsupported conclusions include EPA's determination that lead in soil contributed significantly to house dust lead, additional reductions in blood lead relative to non-abated were observed at 22 months post-abatement, the reductions of lead in house dust persisted for at least 2 years, and the decline in blood lead following abatement continued for at least two years. Perhaps the most egregious error in EPA's analyses is EPA's conclusion that soil abatement caused blood levels to be reduced by 1.86 μg/dL. In fact, this reduction in blood levels was the excess reduction in a group that received soil and dust abatement along with paint stabilization, over that in a group that received paint stabilization only. Thus, an unknown fraction of this reduction may be due to dust abatement rather than soil abatement.

The EPA repeated measures ANOVA applied to Phase I data appears to be a reliable analysis (EPA, 1996, page 5-80), although EPA's description of this model does not conform to the SAS program that implemented the model. This analysis confirmed the analysis by Weitzman *et al.* (1993) that showed a modest, but statistically significant decline in blood lead levels in a group after soil and dust abatement and paint stabilization, compared to a group that experienced dust abatement and paint stabilization. However, EPA considered only in a cursory fashion the extent to which confounding could have played a role in this finding.

EPA's Analysis of Phase II Data of the Boston Study: In EPA's analysis of Phase II of the Boston Study, Group S was used as the control group. As EPA (1996, page 5-82) states, "Group [S] received no further abatement and was an appropriate reference group for the Phase II comparison." To the contrary, about one-half of the Group S homes experienced paint deleading during Phase II (Aschengrau *et al.*, 1994). In fact, the homes of a significant percentage of children in all three groups received paint deleading at this time. According to Aschengrau *et al.*, lead paint removal was offered to all participants, but fewer than one-half accepted the offer. Apparently, all children were included in EPA's analysis of Phase II data, without any control for paint deleading. Although (as described elsewhere in this report) there are a number of other problems with the analysis by Aschengrau *et al.* (1994), at least they were careful to omit from their analysis children whose homes had experienced paint deleading. However, the EPA (1996) report did not even refer to the Phase II paint deleading. As noted in the discussion of the Aschengrau *et al.* (1994) report, even with the best of analyses, it is not possible to draw statistically valid conclusions about the effect of soil abatement from Phase II, due to the lack of an appropriate control group. However, this problem is compounded by EPA's failure to recognize the importance of paint deleading in their analysis. Thus, it would be inappropriate to draw any conclusions from EPA's analysis of Phase II data regarding the effect of soil abatement.

EPA's claim (EPA, 1996, pages 1-19, 6-4) that, in the Boston Study, "Soil was clearly a part of the exposure pathway to the child, contributing significantly to house dust lead.": It is not clear whether the term "significantly" refers to statistical significance or indicates that a substantial fraction of house dust lead comes from soil. Although this conclusion appears twice in the document, there is no discussion of this conclusion anywhere in the document, and there is no reference to an analysis that could serve as the basis for this claim. EPA conducted two types of analyses that could possibly support this claim. The most direct approach analyzes baseline data to determine whether house dust lead is

associated with soil lead. Analyses of this type are found in EPA's pre-abatement cross-sectional structural equation models (EPA Table 5-1) and in EPA's longitudinal structural equation models (EPA Tables 5-30 and 5-32). The most straightforward and perhaps most definitive such analysis — an ordinary regression analysis using soil lead as the dependent variable and house dust as the independent variable — was not performed. In addition, EPA's Table 4.2 presents summary information on median soil and dust levels in various rounds that can be used to compare changes in house dust levels in homes with and without soil abatement. Each of these sources of information are reviewed below.

EPA's Table 5-1 contains results of structural equation models that include imbedded regressions of Round 1 household dust lead levels on soil lead concentrations. Results from 12 models are presented — six that correlate soil lead concentration with lead house dust concentration and six that correlate soil lead concentration with lead house dust load. Although all 12 of these models found a positive association between soil lead and dust lead, this association was statistically significant in only two models, neither of which included window dust. As pointed out by EPA, "pre-abatement soil lead was never a significant predictor of dust lead when window lead was included in the model." Moreover in an apparent reference to these analyses, EPA (1996, page 5-139) stated that "In the pre-abatement model (Round 1) soil lead concentration had little relationship to blood lead or dust lead." This statement appears to contradict the aforementioned conclusion.

EPA's Tables 5-30 and 5-32 also contain results of analyses that regress Round 1 (pre-abatement) blood lead on Round 1 soil lead and household lead concentrations. In Table 5-30, all five of the models found a negative association between pre-abatement soil lead and household dust lead, three of which were what EPA termed marginally significant. In the alternative models considered by EPA that used a "fixed blood lead persistence factor" (EPA, 1996, Table 5.32), all six models showed negative correlations between soil lead and dust lead, four of which were, according to EPA's criterion, marginally significant.

EPA (1996, page 5-127) states, in an apparent reference to Models 17 and 30 in Table 5-32, that there are "significant (possibly different) relationships between dust lead and soil lead, so that there is again evidence of the operation of soil lead abatement by a soil lead to dust lead ... pathway in the Boston study." It is not clear what EPA means by "again evidence" since this is the only place we have identified in the document where EPA attempts to tie this claim to an actual data analysis. It is not completely clear to what analyses EPA is referring in this comment. EPA must be referring to the relation between soil and dust in Round 3, since, as noted above, all six models in Table 5-32 showed a negative correlation between soil-lead and dust lead in Round 1. In Model 30, in which different Round 3 soil-to-dust coefficients are estimated for each group, none of these coefficients are significant. However, the coefficient is positive and statistically significant for Model 17 — as well as for Models 1, 2, and 10 — all of which estimate a common soil-to-dust coefficient applicable to all groups. However, the fact that these coefficients are positive could well have nothing at all to do with the soil-to-dust pathway, since the soil effect upon dust is confounded with treatment in Round 3 data. Note that Group S experienced soil, dust and paint interventions, and consequently even if all dust lead came from interior paint, these interventions would tend to produce both low soil lead values and low dust concentrations in this group. On the other hand, both dust lead and soil lead would remain higher in Group B because neither dust nor soil was abated in this group. Consequently, these positive coefficients could have nothing to do with a soil-to-dust pathway, but could have been produced artificially by intervention.

EPA's treatment of the soil-to-dust pathway issue, as discussed in the previous paragraph, provides an example of EPA's highly selective treatment of data in this report. In interpreting its analyses, EPA ignored the negative soil-to-dust coefficients that it estimated from Round 1 data (where it would be appropriate to reach conclusions regarding the extent to which soil lead influences dust lead), and concentrated instead on positive coefficients obtained from Round 3 data, in which the soil effect on dust is confounded with treatment.

EPA's Table 4-1 contains median values for floor dust lead concentration ($\mu g/g$), floor lead dust load ($\mu g/m^2$), window dust lead concentration ($\mu g/g$), and window lead dust load ($\mu g/m^2$) for Round 1 and Round 3. Of these four measures of house dust, only one (dust floor lead concentration) shows a greater reduction from Round 1 to Round 3 in the soil abatement group (Group S) than in the control group (Group A). Even in this case, the reduction between Round 1 and Round 3 is marginal — only 11%. On the other hand, the dust reduction between Round 1 and Round 3 is larger in Group A than in Group S for median floor dust lead load (by 14%) and for median window dust lead load (by 18%). In addition, the median window dust lead concentration in Group S increased by 7885 $\mu g/g$ (60%) between Rounds 1 and 3, and decreased by 4017 $\mu g/g$ (20%) in Group A. Similar results are obtained if Round 1 is compared to Round 2. Thus, there is essentially no evidence from Table 4-1 that soil lead abatement reduced either median lead dust concentration or median lead dust load.

Thus, EPA's analyses do not support EPA's contention that soil lead contributed significantly to house dust lead. EPA concluded elsewhere in the document (EPA 1996, page 5-139) that pre-abatement soil lead concentration "had little relationship to blood lead or dust lead"; this appears to be a more reasonable conclusion.

EPA's claim (pages 1-19 and 6-4) that, in the Boston Study, "reductions of lead in both soil and house dust persisted for at least two years.": This claim appears to be valid regarding lead in soil. On the surface, it is also true regarding lead in house dust. However, EPA appears to be implying that the reduction in dust lead is due to soil abatement, whereas there is no evidence to support this in their report. Group S is the only group that was observed for 2 years following soil abatement. As EPA's Table 4-2 indicates, Group S post-abatement median floor dust concentrations remained reduced in each of Rounds 2, 3, and 4. However, almost the exactly same pattern was observed in both Groups A and B, neither of which experienced soil abatement until Round 4. (Moreover, one must be very

cautious when interpreting Round 4 data, since an appreciable number of the residences experienced paint deleading between Rounds 3 and 4.) Consequently, the reduction in lead in house dust cannot be attributed to soil abatement, and is more likely to be due to lead paint intervention.

EPA's claim (pages 1-31 and 6-19) that, in the Boston Study, "[blood lead] reduction continued to increase for two years following abatement": This statement appears to indicate that, not only did blood lead reductions continue for 2 years following abatement, the rate of reduction continued to increase over this period. Since Group S was the only group in the Boston Study that was observed for 2 years following soil abatement, this claim must be based on this group.

EPA's Table 4.2 indicates that median blood lead in Group S was 10 μg/dL in each of Round 2 (7 months post-abatement), Round 3 (11 months post-abatement) and Round 4 (22 months post-abatement). Lack of any change in blood lead in this group between Rounds 2 and 4 are also supported by the geometric means (EPA, 1996, Table 4.2) and EPA's graph of blood lead concentrations (EPA, 1996, Figure 5-34). Thus, blood lead in Group S did not appear to change at all between 7 months and 22 months post-abatement. There appears to be no basis for EPA's claim to the contrary anywhere in their report.

This claim appears to be related to EPA's "key finding" from the Boston study (EPA, 1996, pages 1-19 and 6-4) that "Additional reductions in blood lead of about 2.0 μg/dL (relative to non-abated) were observed at 22 months post-abatement for children in houses where the soil lead was abated and the interior house dust lead was consequently reduced and remained low." It is not clear to what the phrase "relative to non-abated" refers, since by 22 months the soil of all the properties in the study had been abated. Further, EPA presents no analysis of blood lead 22 months post-abatement in children from houses where "dust lead was reduced and remained low," and, as noted above, median blood lead in Group S children did not change between Rounds 2 and 4 (7 and 22 months post-abatement). According to Aschengrau *et al.* (1994), only 14 children had their

blood lead monitored for 22 months following soil abatement, without additional interventions.

EPA's Repeated Measures Analysis of Variance Analysis: EPA's repeated measures analyses simultaneously modelled pre- and post-abatement blood levels while accounting for the dependence between the pre- and post-abatement blood lead levels in an individual child. Thus this method complements the analysis of Weitzman *et al.*, who modelled post-abatement levels, using the pre-abatement level as a covariate in the model. The repeated measures analysis also accounts for the dependence between blood levels of children in the same household, for which the analysis by Weitzman *et al.* did not account.

EPA (1996, page 5-16) described their repeated measures model as follows:

$$Y_{ir} = G_{gr} + H_{h(g)} + I_{i(gh)} + \epsilon_{ir},$$

where Y_{ir} is the blood level of the ith child in the rth round, G_{gr} is a "group effect term" (i.e., an indicator variable that specifies membership in abatement Group S, A, or B), $H_{h(g)}$ is a "household effect term," $I_{i(gh)}$ is an "individual child effect term", and ϵ_{ir} is an "error term." However, after reviewing the SAS program used by EPA to implement their repeated measures analysis, it appears that a better description of their model is the following:

$$Y_{ir} = G_{gr} + \delta_F + \epsilon_{ir},$$

where Y_{ir} and G_{gr} are as described above, δ_F is a random effect that is constant within families (which causes the blood lead levels of siblings to be dependent), and ϵ_{ir} is a random error term, with such terms being independent across children, but dependent across rounds within children [$Cov(\epsilon_{i1}, \epsilon_{i2}) > 0$], which causes a child's blood lead levels in Rounds 1 and 2 to be dependent.

One potential limitation of this analysis is that, while dependence among the blood leads of children within a single home is accounted for, dependence of blood lead levels of children living at the same premises is not accounted for. There were a total of 150 children in 125 households, but only 100 premises in the three study groups combined. Consequently, it might be important to take into account the dependence of blood leads among children residing at the same premise.

Using this model, EPA found a statistically significant difference in blood lead between Group S and Group B ($p = 0.004$) and between Group S and Group A ($p = 0.016$). The estimated differences in blood lead levels were reasonably close to those estimated by Weitzman *et al.* (1993): 1.87 μg/dL (vs. 1.49 obtained in Weitzman *et al.* base model) comparing Group S and Group B, and 1.54 μg/dL (vs. 1.28 obtained in Weitzman *et al.* base model) comparing Group S and Group A. However, as noted earlier, differences between Groups S and B cannot be attributed to soil remediation.

EPA's repeated measures analysis of covariance: EPA's repeated measures analysis of covariance conducted by EPA on Phase I Boston data are reported in Tables 5-18 and 5-19. Apparently, the models used in these analyses are the same as used in EPA's repeated measures analysis of variance (Table 5-5) except that log dust lead was added as a covariate. EPA does not state whether the dust lead data used in this analysis was from Round 1 or Round 3. Apparently, it was Round 3 data, although is not clear why EPA wanted to include Round 3 dust lead as a covariate, since intervention may have affected dust lead, and therefore any effect of treatment on blood lead may be underestimated by inclusion of this variable. Indeed, a comparison of Tables 5-5 and 5-18 shows that whatever dust variable that was included caused the treatment effect to change dramatically, assuming that the first three rows of Table 5-18 contain values of the treatment effect. (Although these rows are not labeled, it is difficult to think of anything else to which they could refer.) If this interpretation is correct, then Table 5-18

indicates that soil abatement caused blood lead levels to increase (based on comparison of Groups S and A), once dust was controlled for.

EPA's description of its repeated measures analysis of covariance proved impossible to follow. A table referred to in the text containing results of using log dust loading as a covariate instead of log dust concentration apparently was omitted. In addition, there are unexplained gaps in Tables 5-19 and 5-20. We could not make sense of EPA's description of Table 5-18. EPA's comment that "the increasing slope of the log blood lead versus log dust lead concentration grew much more strongly in the dust abatement group [A] between Round 1 and Round 3 than in either the soil abatement group [S] or the control group [B]" seems to refer to values of the group effect variable (based on the similarity between Tables 5-18 and 5-5 — the values are not explained in Table 5-18 — and since the lower part of the table contains the coefficients for log dust concentration), and, if so, does not refer to the dust concentration coefficient at all.

EPA's longitudinal structural equation analysis: EPA used structural equation modelling to attempt to model simultaneously the relationships among pre-abatement and post-abatement (Round 3) values of blood lead, soil lead, floor dust lead, and window dust lead. EPA describes this work as an "exploratory model building activity that does not readily lend itself to hypothesis testing" (EPA, page 5-68). After review of this modelling effort, we agree with this conclusion. However, despite this caveat, a quantitative estimate of the effect of soil abatement upon blood lead is prominently cited in the summary of the EPA results from the Boston data (EPA's Tables 1.4 and 6.1).

In an effort to better understand EPA's structural equation analyses, we requested and received, courtesy of Dr. Robert Elias, a copy of a representative SAS program that was used to conduct these analyses. These programs revealed some important features of these analyses that were not provided in EPA (1996).

Apparently, in order to get some of the model runs to converge, EPA appended a penalty function to the objective function of some (but not all) of the

models. The penalty function had the effect of tending to make the coefficient for Round 1 blood lead in modeling Round 3 blood lead to assume a value near 0.5. When we ran Model 1 (EPA, 1996, Table 30) without this penalty function, it did not converge, whereas when the penalty function was included, it converged to the value reported by EPA.

EPA described fitting of 32 structural equation models, but presents results for only six. EPA claimed that these six were the "best-fitting," but did not explain what criteria were used to evaluate fit. Based on our experience running the programs provided to us by EPA, it seems likely that the SAS program did not converge for a number of these models. If this was the case, it is incorrect to describe the models for which the program converged as "best-fitting." Many of the models applied by EPA were not pertinent to the central question of the study, regarding whether soil abatement resulted in reduced blood lead levels. Of the six structural equation models for which quantitative data are reported (EPA, 1996, Tables 5-29 to 5-38), only two estimated the effect of soil abatement.

The longitudinal structural equation models used by EPA included floor dust lead, soil lead, and window dust lead from Rounds 1 and 3 and blood lead from Round 1. However, the analyses most emphasized by EPA (EPA, 1996, Tables 5-30 and 5-32) did not control for a number of factors that could influence blood lead levels, including, among others, gender, age, race and ethnicity, and socioeconomic status. Gender was considered only by analyses that involved only males or only females. Age was considered in only one analysis, which considered only children aged 18–41 months. None of the structural equation analyses controlled for race or ethnicity, despite the fact that Weitzman *et al.* (1993) found that soil abatement was not a significant predictor of blood lead when race/ethnicity was included in their analysis. Also, none of EPA's structural equation analyses controlled for socioeconomic status, nor did they control for any of the 19 other potential confounding variables studied by Weitzman *et al.*

Further, some of the variables that were included in EPA's structural equation model may not have been controlled for adequately. For example,

Weitzman *et al.* (1993) found, not surprisingly, that Round 1 blood lead was a highly significant ($p < 0.0001$) predictor of Round 3 blood lead. However, in EPA's structural equation analyses Round 1 blood lead was either an insignificant or only a marginally significant predictor of Round 3 blood lead (EPA, 1996, Table 5-30). The reason for this may be related to the fact that in EPA's structural equation models it is assumed that Round 1 blood lead can be adequately predicted by Round 1 soil and dust levels. This assumption was not valid, as not a single structural equation model applied by EPA to the complete cohort (EPA, 1996, Tables 5-30 and 5-32) found a statistically significant relationship between soil lead and blood lead in Round 1. This misspecification may have caused this analysis not to find the association between Round 1 and Round 3 blood lead. Thus, it appears that even though Round 1 blood lead was included in EPA's structural equation models, Round 1 blood lead may not have been controlled for adequately.

The main conclusion drawn by EPA from its structural equation modelling effort was its estimates of the reduction in blood lead from soil abatement (1.86 μg/dL obtained by comparing Group S to Group B and 1.56 μg/dL obtained by applying Group S to Group A), as these values are included in both EPA's summary Table 1-4 and Table 6-1. As noted earlier, only differences between Group S and Group A can be ascribed to soil abatement. Neither of these values were statistically significant (page 5-126); yet they were used as primary support for EPA's claim for an effect of soil abatement in reducing blood lead levels.

EPA utilized the SAS procedure MODEL to estimate the parameters in their structural equation models. SAS provides a number of procedures for estimating parameters of such models. To determine how the answer obtained by EPA would differ using different estimation procedures, we fit EPA's "Model 17" with the blood lead persistence factor fixed (EPA, 1996, Table 5-32; this is the model fit that was emphasized by EPA in summary Tables 1-4 and 6-1) using four additional estimation procedures provided by SAS. The estimation procedure used by EPA estimated that soil abatement resulted in a reduction in blood lead of 1.56 μg/dL

(based on comparisons between Group S and Group A). The five estimation procedures (using SAS nomenclature, ITGMM [used by EPA], FIML, GMM, IT3SL, and N3SLS) produced an array of estimates ranging from 0.58 to 1.93 μg/dL. Thus, the results of EPA's structural equation modelling appears to be quite sensitive to the specific estimation procedure used by EPA. There appears to be no compelling reason to prefer one estimation procedure over the others.

In the application of EPA's longitudinal structural equation Model 17, which was used to calculate the size effect of 1.86 reported in EPA's summary Tables 1-4 and 6-1, the effect of Round 1 blood lead was not estimated in the conventional fashion from fitting the model to the data. Instead, this effect was fixed at the value obtained from fits of two different models. EPA did not explain the reason for this highly unorthodox approach, stating only that "Since the persistent effect of pre-abatement blood lead is the largest single component of post-abatement blood lead for most of the children in the sample,[1] we decided to also evaluate models in which this regression coefficient was held fixed at the Model 10 and 11 optimal value of 0.589." It is not at all clear why EPA thought Model 17 could be improved by replacing this parameter estimate by a parameter value obtained from different models. It is even less clear why EPA selected Models 10 and 11, since neither of these models addressed the effect of soil abatement. However, fixing this parameter in this manner caused the model estimates to be substantially different (compare EPA, 1996, Tables 5-30 and 5-32). Thus, it appears that the effect estimates of 1.86 and 1.56 obtained by EPA are very sensitive to the value assumed for this parameter. Although EPA's decision to rely upon an analysis in which this parameter was fixed appears to be an arbitrary one, it had a critical effect upon the magnitudes of the effect sizes emphasized in their conclusions. The only plausible reason we can think of as to

[1]However, this cannot be concluded from this analysis, since, as noted earlier, pre-abatement blood lead was not a significant predictor post-abatement blood lead in EPA's structural equation models.

why EPA emphasized this particular analysis is that it provided effect estimates that were similar to those obtained by Weitzman *et al.* (1993) and by EPA in its repeated measures analysis. However, this similarity was probably serendipitous.

Another disturbing feature of these analyses is that model estimates for Models 10 and 11 were substantially different in EPA's Table 5-32 from those in EPA's Table 5-30 even though the only difference in these analyses is that in Table 5-32 the Round 1 blood lead parameter in these models was fixed at the value obtained in the unconstrained fit reported in Table 5-30. Theoretically, fixing this parameter in this way should have no effect upon the numerical values of other parameter estimates. The fact that estimates for Models 10 and 11 are substantially different in Tables 5-30 and 5-32 strongly suggests that the algorithm used to make the calculations did not converge properly. If this is the case, the reported estimates are likely to be meaningless. In fact, we determined from our own analyses that a penalty function was likely added to the objective function in the results for Model 17 reported in Table 5-30 but not in results reported in Table 5-32. This use of a penalty function was not reported by EPA. It also appears that EPA's reliance upon an analysis that involved such penalty functions was arbitrary, since there is no apparent scientific basis for accepting a model that incorporated a penalty function over one that did not.

In summary, we found numerous shortcomings with EPA's structural equation modelling. These models tended to be unstable, and results reported by EPA were strongly influenced by the estimation procedure used by EPA, and whether EPA included a penalty function in the estimation procedure. The quantitative estimate of the effect of soil abatement emphasized by EPA relied upon the *ad hoc* approach of fixing one parameter at the value estimated by another model that did not even include a term for the soil abatement effect. These analyses did not attempt to control for a number of potentially important confounding variables, including socioeconomic status and ethnicity. EPA stated that the purpose of structural equation modelling was "to elucidate pathways for environmental lead exposure from source to child" and "is an exploratory model-

building activity that does not readily lend itself to hypothesis testing." Based on our review of this work, we question whether even this limited goal was achieved.

Miscellaneous comments on the EPA (1996) integrated report: None of the earlier reports of the three individual studies (EPA, 1993a,b,c) are in the EPA (1996) reference list, although EPA quotes from these documents in several places. Additionally, the published report of Phase I of the Boston study (Weitzman *et al.*, 1993) is inexplicably absent from the reference list and is never referred to in the EPA report, although the published report of the follow-up study (Aschengrau *et al.*, 1994) is referenced and quoted.

(EPA, 1996, page 1-13, paragraph 2) "a significant effect of intervention for both the BOS PI-S [Group A] and BOS-SPI [Group S] groups"

Presumably the control group for these findings was Group B in each case. However, Group A and Group B blood lead levels changes did not differ significantly.

(EPA, 1996, page 5-38) "the window data indicate that both paint and soil sources of lead were interrupted, at least temporarily."

The basis for this conclusion is not apparent from the graphs from which it is derived. There appears to be essentially no changes in lead window dust in any group. Similarly, the basis for the conclusion on page 5-36, regarding window dust, that "paint stabilization and soil abatement appear to have been effective and persistent for several hundred days, similar to floor dust" is unclear. In fact, window dust lead concentrations were higher 11 months post-abatement than pre-abatement levels in the main treatment group (Group S).

(EPA, 1996, page 1-21) "The Afro-American children seemed to show larger responses to dust abatement than did the sample as a whole."

This apparently refers to the log dust regression coefficients (or differences in these regression coefficients. It is unclear from EPA tables exactly what these coefficients represent) in EPA's repeated measures of covariance (Tables 5-18 to 5-24). Actually, one of the coefficients in question is positive while the other two are negative, while all are larger in absolute value when the analysis is restricted to Afro-Americans. Results for different groups, therefore, appear to be conflicting and EPA's conclusion appears to be unwarranted.

EPA (1996, page 2-17) concluded that "under optimal conditions any environmental .. intervention can likely only achieve a 40 to 50% reduction in child blood lead concentrations within a year after abatement."

There is no way to deduce this from information provided by EPA, since EPA provides only a hypothetical graph and no references to support this conclusion.

(EPA, 1996, Table 5-2)

The entries in this table do not correspond to the models as defined in Figure 5-45.

(EPA, 1996, page 6-1) "Larger effects were identified for ... children of Afro-American ancestry, than for the sample as a whole."

It is not clear what "effects" are being commented upon, but it appears that it is a reference to effect of soil abatement. If so, the conclusion is not supported by EPA's analyses. In comparisons between Groups S and A (the appropriate groups to compare when evaluating the effect of soil abatement) categorized by black and non-black (Table 5-7), the effect in non-black was 1.72 μg/dL lead in blood (marginally significant) and only 1.13 μg/dL in blacks (non-significant).

Independent evaluation of the effect of soil abatement on blood lead levels in the Boston Study

As described earlier, it is not possible to draw statistically valid conclusions from Phase II of the Boston Study because of the complicated nature of the pattern of interventions and the lack of a suitable control group. Consequently, the present evaluation is limited to Phase I data only. Moreover, since soil abatement is completely confounded with the dust abatement in comparisons between Group S and B, the present evaluation only considers data from Groups S and A.

Statistical evaluation of the effect of soil abatement upon blood lead in the Boston Study: In this section we present a new statistical evaluation of the evidence for an effect of soil abatement upon blood lead in the Boston Study. This evaluation utilizes randomization tests (Edgington, 1987). A randomization test does not require any assumption about the underlying distribution of blood leads, and consequently is a more generally applicable type of testing procedure than the procedures used by either Weitzman *et al.* (1993) or EPA (1996). A randomization test seems particularly appropriate in this case, since the chief condition required for the validity of a randomization test is that subjects were randomly assigned to treatment and control groups, which is the case in the Boston Study. The randomization test implemented provides an exact test for the null hypothesis that soil abatement had no effect upon blood levels, i.e., that the differences in blood lead levels between Groups S and A are totally a result of the random assignment of premises to Groups S and A. In addition to not requiring any assumptions about how blood leads are distributed, the randomization tests also take into account the fact that the premise was the basic experimental unit in the Boston Study, rather than the child.

The Weitzman *et al.* (1993) analysis of the Phase I data assumed that the blood lead data are normally distributed. Similarly, all of the analyses conducted by EPA assume that blood leads are distributed either normally or lognormally. However, the density of Round 1 blood leads decreases steadily with increasing

blood lead levels above the cutoff of 7 μg/dL, and consequently neither of these distributions are likely to adequately describe the blood lead data. Moreover, neither the Weitzman *et al.* (1993) analysis nor any of the EPA (1996) analyses accounted for the fact that the premise formed the basic experimental unit in the Boston Study.

The randomization tests were performed as follows. First, a conventional regression analysis of post-abatement (Round 3) blood lead was applied to the data in Groups S and A, using as explanatory variables a category variable for group assignment, the pre-abatement (Round 1) blood lead level (base case), and various additional explanatory variables (gender, race/ethnicity, age, and socioeconomic status). The value of the t-statistic for a standard test of a group assignment effect was recorded. Then premises in Groups S and A were randomly reassigned to Groups S and A, with the number of premises in each group remaining fixed at the original values. This randomization procedure was repeated 1000 times to produce 1000 sets of randomized data. The corresponding t-statistic was computed for each set of randomized data. The proportion of the 1000 sets of randomized data for which the t-statistic equalled or exceeded the t-statistic from the actual data was then computed. This proportion represents the one-sided p-value for a test of the hypothesis that post-abatement blood lead levels were lower in Group S than in Group A. Note that, in addition to being non-parametric (i.e., requiring no assumptions about the distribution of blood lead values), this procedure accounts for the fact that premises, rather than children, were randomly assigned to treatment groups and consequently premises represent the basic experimental unit.

Table 1 gives the results of this randomization analysis. The first column of this table gives estimates of the treatment effect, i.e., the average reduction in blood leads between Rounds 1 and 3 in Group S minus the corresponding reduction in Group A. The second column gives the randomization one-sided p-values for the significance of the treatment effect. The third column gives the one-sided p-values based upon the t-statistic and the normal distribution.

Table 1 shows that there was a statistically significant ($p < 0.05$) effect of abatement upon blood lead levels. This effect is statistically significant in the base model, and also when age, gender, race and socioeconomic status (SES) are controlled for, either individually or jointly. Randomization p-values and the standard p-values based upon the assumption of a normal distribution agree closely in each case. Whereas Weitzman *et al.* (1993) found that the treatment effect was not significant when race and ethnicity were controlled for, Table 1 shows that, when the analysis involves only Groups S and A, the effect of soil abatement remains significant even when these two factors were included in the model.

Estimation of the size of the reduction in blood lead resulting from soil abatement: Weitzman *et al.* (1993) estimated the effect of soil abatement upon blood lead to be 1.28 μg/dL, based on an analysis that adjusted for pre-abatement blood lead only. However, the estimate is likely to be improved by adjusting for other variables that are known to be associated with blood lead. In their analysis that also adjusted for race/ethnicity, Weitzman *et al.* estimated the effect of soil abatement to be 0.92 μg/dL. Based on our randomization analysis, we estimated the effect of soil abatement to be in the range of 1.15 to 1.37 μg/dL, based on estimates that adjust for pre-abatement blood lead levels, age, gender, race/ethnicity, and SES. When all of these variables were controlled jointly, the value of 1.37 μg/dL was obtained.

Care must be taken when extrapolating an estimate of the effect of soil abatement in the Boston Study to other situations and localities, as there are a number of reasons why this approach would tend to overestimate the effect of soil abatement in general. First, the median soil lead level in Boston was in excess of 2000 ppm, which appears to be higher than in most urban settings. Abatement in urban areas with lower soil lead levels would likely result in smaller reductions in blood lead levels than were observed in the Boston Study. This is borne out by the Cincinnati and Baltimore studies, where soil lead levels were lower and no benefit of soil abatement in reducing blood lead levels was identified. Second, only

children whose blood lead levels were between 7 and 24 μg/dL were included in the Boston Study. Although Weitzman *et al.* do not provide information on the blood lead levels in children who were excluded because their blood lead levels were not in this range, based on the distribution of blood lead levels in the study, it appears that the vast majority of excluded children had blood lead levels below 7 μg/dL. The reduction in blood lead resulting from soil abatement would very likely be lower in such children. Third, all of the analyses of the Boston data discussed herein omitted two children from Group S whose blood lead levels increased considerably between pre-abatement and post-abatement (from 19 to 35 μg/dL and 12 to 43 μg/dL). Weitzman *et al.* indicated that the increases in blood lead in these two children were likely due to exposure to leaded paint at another site. However, exposures such as this occur in real life situations, and we do not know if such cases are over- or under-represented within this cohort. If these two children are included in the analysis, we estimated the effect of abatement as –0.17 μg/dL; i.e., if the two outliers are included, there is an estimated net increase in blood in the soil abatement group (Group S) relative to the control group (Group A). Fourth, the reductions in blood lead observed in the Boston Study were in addition to effects of paint stabilization and dust abatement. Consequently, the Boston Study does not address the efficacy of soil abatement by itself, but only in the presence of these other interventions.

References

Aschengrau A, Beiser A, Bellinger D, Copenhafer D, Weitzman M. (1994). The impact of soil lead abatement on urban children's blood lead levels: phase II results from the Boston lead-in-soil demonstration project. *Environmental Research* 67, 125-148.

Edgington ES. (1987) *Randomization Tests*. Second Edition. Marcel Dekker, Inc. New York. 341 pp.

EPA (1993a) Urban Soil Lead Abatement Demonstration Project. Volume II: Parts 1 and 2, Boston Report. United States Environmental Protection Agency, EPA/600/AP-93/001b.

EPA (1993b) Urban Soil Lead Abatement Demonstration Project. Volume III: Parts 1 and 2, Baltimore Report. United States Environmental Protection Agency, EPA/600/AP-93/001c.

EPA (1993c) Urban Soil Lead Abatement Demonstration Project. Volume IV, Cincinnati Report. United States Environmental Protection Agency, EPA/600/AP-93/001d.

EPA (1996) Urban Soil Lead Abatement Demonstration Project. Volume I: EPA Integrated Report. United States Environmental Protection Agency, EPA/600/P-93/001aF.

Pirkle JL, Brody DJ, Gunter, EW, Kramer RA, Paschal DC, Flegal KM, Matte MD. (1994) The decline in blood lead levels in the United States. The National Health and Nutrition Examination Surveys (NHANES). *JAMA* 272: 284-291.

Weitzman M, Aschengrau A, Bellinger D, Jones R, Hamlin JS, Beiser, A. (1993). Lead-contaminated soil abatement and urban children's blood lead levels. *JAMA* 269, No. 13, 1647-1654.

Table 1

Results of Randomization Tests for Effect of Soil Abatement Upon Blood Lead (Group S compared to Group A)

	Effect[a] (μg/dL)	Randomization p-value[b]	t-statistic P-value[b]
Base model	1.30	0.009	0.009
+ age	1.34	0.012	0.008
+ gender	1.34	0.010	0.005
+ race and ethnicity	1.13	0.028	0.026
+ SES	1.22	0.016	0.013
+ gender & race and ethnicity & age & SES	1.37	0.010	0.011

[a] Blood lead reduction in Group S between Rounds 1 and 3 minus corresponding blood lead reduction in Group A. Round 1 blood lead was included in each model.

[b] P-value for one-sided test.